Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2003 Proceedings

International Conference on Information Systems (ICIS)

December 2003

Representing Things and Properties in Conceptual Modeling: Understanding the Impact of Task Type

Graeme Shanks
Monash University

Jasmina Nuredini
Monash University

Daniel Tobin
Monash University

Ron Weber
The University of Queensland

REPRESENTING THINGS AND PROPERTIES IN CONCEPTUAL MODELING: UNDERSTANDING THE IMPACT OF TASK TYPE

Graeme Shanks
Monash University
Clayton, Victoria Australia
graeme.shanks@infotech.monash.edu.au

Jasmina Nuredini
Monash University
Clayton, Victoria Australia
jasmina.nuredini@infotech.monash.edu.au

Daniel Tobin
Monash University
Clayton, Victoria Australia

Ron Weber
The University of Queensland
Brisbane, Queensland Australia
weber@commerce.uq.edu.au

Abstract

The representation of things and properties is a fundamental issue in conceptual modeling. The proponents of different modeling approaches, for example entity relationship modeling and object-role modeling, offer very different advice about the distinction between things and properties and their representation. We use ontological theory to provide guidelines about how things and properties should be represented. Previous experimental work has provided evidence to support the use of ontologically sound representations of things and properties in conceptual modeling. However the results also indicate that the type of task undertaken (for example comprehension, problem solving, discrepancy checking, and decomposition) may also impact the use of conceptual models. In this paper, a research project is proposed to examine the sorts of tasks that are best supported by distinguishing between things and properties in conceptual modeling.

Keywords: Conceptual modeling; information systems development; ontology; entity; thing; object; attribute; property; entity-relationship model; normalization

Introduction

The notions of things and properties and their representation in conceptual models are problematic for a number of reasons. First, not all conceptual modelers agree that things and properties are distinct phenomena. For example, object-role modelers (e.g., Halpin 2001) subscribe to the views of nominalist philosophers who "dispense with properties…and reduce everything to things, their names and collections of such" (Bunge 1977, p 57). Second, those who argue that things should be distinguished from properties of things have difficulty showing how the distinction ought to be made—a difficulty also faced by philosophers who argue it should be sustained (e.g., Denkel 1996). Third, further confusion about the representation of things and properties arises when database design considerations are taken into account. For example, Simsion and Witt (2001, p 104) argue that attributes in a conceptual model simply correspond to columns in a relational model. In other words, database design considerations indicate which phenomena in the world ought to be considered as properties.

We argue that the question of whether things and properties should be distinguished in conceptual models and, if so, that how they should be distinguished ought to be based on sound underlying theory about the structure and dynamics of phenomena in the world. This will enable conceptual models to better support the discovery and documentation of user views of a domain. As a result, the models will provide a better basis for informed discernment, reconciliation, and compromise among stakeholders during the development and implementation of an information system (Hirschheim et al. 1995). In this light, we have undertaken a number of theoretically driven empirical evaluations of alternative conceptual modeling representations of things and properties. Overall, our research has had a threefold motivation:

- Improvements in conceptual modeling practice potentially have high payoffs because they enable errors to be resolved early in the systems development process, thereby mitigating the high costs associated with late detection of errors (Moody and Shanks 1998).

- Conceptual models can be used to evaluate the fit between the business models embedded in an enterprise application software package and an organization's requirements (Sia and Soh 2002). Preparing high-quality conceptual models is critical to an accurate assessment of fit.

- We are seeking to improve conceptual modeling practice. If things ought to be distinguished from properties of things in a conceptual model, we seek to provide clear guidance to practitioners on how the distinction should be made.

Based on our prior work, we now believe that certain types of tasks undertaken by users of conceptual models are best supported when things and properties are distinguished in the models. In the research described in this paper, we are seeking empirical evidence to support our view. In the next section, we discuss the theory that underpins our empirical work. The third section provides an overview of the results of our prior research, discusses these results in the context of other related research, and defines the objectives of the empirical study we describe herein. The fourth section discusses the research method we used. The fifth section discusses our preliminary findings. Finally, we present some limitations of our research.

Theoretical Basis of the Study

Two bodies of theory underpin our work. The first is Bunge's (1977) ontological theory. This theory has proved useful in prior research on conceptual modeling that has examined the strengths and weaknesses of various conceptual modeling practices (e.g., Bodart et al. 2001, Shanks et al. 2003).

Bunge's theory simply postulates that things and properties of things are two separate types of phenomena in the world. Nonetheless, the distinction allows further ontological constructs to be derived—for instance, the state of a thing, an event in a thing, couplings between things, and systems. Our view is that these constructs accord with widely held beliefs about how the world is structured. In the absence of the distinction, however, constructs like the state of thing have no meaning. Objects in the world simply have a value, which enables only limited representation of the world. For instance, there is no apparent basis for clustering object values into a vector to describe phenomena like customers or stock items.

The second body of theory that underpins our research is cognition theory. Extensive research has shown that humans cognitively cluster phenomena that they perceive to be related (e.g., Bousfield 1953). Clustering appears to provide a means for humans to deal with the complexity they often encounter in their perceptual worlds (Miller 1956). By focusing on clusters, they reduce cognitive load and enhance their abilities to understand the world. Properties of things naturally cluster with the things to which they belong. Perceiving the world in terms of things and their properties, therefore, helps humans to mitigate the cognitive problems they experience when they perceive phenomena to be complex.

Previous Empirical Work

In previous empirical work, we undertook a laboratory experiment with 80 end users of information systems to test whether alternative representations of phenomena that might be classified as things and properties of things impacted their performance in relation to various tasks they had to perform with conceptual models (Shanks et al. 2003).

We used four alternative representations of the phenomena. The first, which we termed the *ontologically sound* level, represented things as entity types and properties as attribute types in an entity-relationship (ER) diagram. The second, which we termed the *partially ontologically sound* level, represented only *mutual* properties (properties of n-tuples of things) as entity types. *Intrinsic* properties (properties inherent to an individual thing) were still represented as attribute types. The third, which we termed the *normalized* level, represented mutual properties and some intrinsic properties as entity types. This level complied with the approach to representing application domains via ER diagrams used by many practitioners (Simsion and Witt 2001). The fourth, which we termed the *entity only* level, represented both things and properties as entity types. This level follows the principles used by object-role modelers (Halpin 2001).

The dependent variable, performance, was evaluated in three ways: comprehension performance, problem-solving performance, and discrepancy-checking performance. *Comprehension* involved the end users using a conceptual model to understand the surface-level features of the domain represented by the model. *Problem solving* involved the end users using a conceptual model to solve problems that might arise in the domain. *Discrepancy checking* involved end users comparing a conceptual model against a text to evaluate whether the conceptual model represented the semantics manifested in the text accurately and completely.

In light of results obtained by Gemino (1999), Bodart et al. (2001), and Shanks et al. (2002), we expected that the different representations primarily would affect end users' deep-level understanding of a domain and thus their problem-solving performance. To our surprise, our results showed that the type of representation affected only comprehension performance. After careful consideration of our results, we now believe that humans differentiate between certain features of the world because it affects their ability to perform specific tasks. In other words, certain types of differentiation facilitate some tasks but not others. In this research, therefore, our specific objective is to try to identify the kinds of cognitive activity (and ultimately the sorts of uses of conceptual models) that will be facilitated by distinguishing between things and properties of things in conceptual models.

Research Method

We are using a cognitive process-tracing research approach to better understand the cognitive behavior of participants during several different types of problem-solving task (Ericsson and Simon 1984). Our current study includes two of the four alternative representations described above—the ontologically sound level and the normalized level—and two of the performance tasks—comprehension and problem solving.

Materials

Four sets of materials were developed for the study. The first was a summary of the ER symbols used in the diagrams provided to participants in the study. The second comprised a personal-profile questionnaire to obtain information about participants' backgrounds. The third comprised two ER diagrams of alternative models of a sales order domain (one that is understood widely), reused from our previous experimental study. The fourth comprised five comprehension questions and five problem-solving questions, again reused from our previous experimental study (although the problem-solving questions were made slightly more complex to yield richer protocol data).

Participants

Participants in the experiment were 12 individuals working in industry. They did not play an information technology role in their organizations, nor did they have information systems/technology qualifications. In essence, they acted as surrogate end users.

Procedures

The materials were first pilot tested with two participants. The materials were found to be fine although regular prompting of participants was found to be necessary to stimulate speaking aloud during the tasks.

Participants were first assigned randomly to one of the two alternative representation groups and within each group the sequence of tasks was altered for every second participant. Participants were then run singly through the study. When they arrived to undertake the experiment, they were asked to complete a consent form and the demographic survey. The "speak aloud" approach to data collection was explained. A camcorder mounted on a tripod was focused on the ER models and videotaped participants as they indicated navigation of the models with a pencil and recorded their verbalizations.

Next participants were given the document that explained the ER modeling symbols. Participants were permitted to discuss the symbols with the researchers until they indicated they felt confident with them. They retained and could refer to the ER modeling symbol summary throughout the study.

When participants indicated they were ready to begin, they were then given either the ontologically sound or normalized ER diagram and asked to work through the first task (either comprehension or problem solving). They were prompted to speak aloud to explain their cognitive behavior if periods of silence occurred. After a brief pause, participants were asked to work through the second task. At the conclusion of the tasks, participants were thanked and dismissed.

Preliminary Findings

Verbal data on all videotapes has been transcribed. A coding scheme has been established using the problem solving literature (for example, Newell and Simon 1972), similar previous studies of data modeling (for example, Batra and Davis 1992), and the content of the transcribed data. The coding scheme comprises five categories.

- *Understanding Question*: includes reading the question, seeking clarification, identifying assumptions and constraints, and recognizing the problem posed
- *Identifying Model Segment*: includes locating appropriate parts of the model and matching them against key concepts in the question
- *Articulating Model Semantics*: includes verifying semantics of symbols in the model and rereading the symbol summary
- *Preparing Solution*: includes developing solutions and simulating and revising solutions against the question
- *Evaluation*: includes selection of alternative answers and developing justifications

The transcribed data was then partitioned into segments and each segment assigned to a predefined category within the coding scheme. In addition, video data was used to help identify start and end times for each segment. Data was coded independently by two of the authors and differences were reconciled.

For the relatively simple comprehension task, participants spent similar amounts of time in understanding, preparing, and evaluating tasks for both the ontologically sound and normalized models. However, participants spent approximately twice as much time identifying model segment and articulating model semantics for the normalized model as the ontologically sound model. This indicates that, for the comprehension task, distinguishing between things and properties facilitates user understanding of the model.

For the relatively complex problem solving task, participants spent significantly more time in all tasks than in the comprehension task, reflecting the differences in complexity of the tasks. Again, participants spent similar amounts of time in understanding, preparing, and evaluating tasks for both the ontologically sound and normalized models. Participants spent approximately 70 percent more time identifying the model segment for the normalized model as the ontologically sound model. However, participants spent approximately 40 percent more time articulating model semantics for the ontologically sound model than for the normalized model. This indicates that although distinguishing between things and properties facilitates locating relevant concepts in the model, task complexity affects articulation of model semantics.

Limitations and Future Research Directions

The major limitation of the study is the laboratory context and materials that are limited in scope and somewhat artificial. Nonetheless, our task has enough realism that our results should be robust in other settings involving thing-property representations. Ongoing research will include further analysis of the data collected and future work involving analysis of cognitive behaviors for other types of task including discrepancy checking and decomposition.

Acknowledgments

An Australian Research Council Discovery Grant funded this research.

References

Batra, D., and Davis, J. "Conceptual Data Modeling in Database Design: Similarities and Differences between Novice and Expert Designers," *International Journal of Man-Machine Studies* (37), 1992, pp. 83-101.

Bodart, F, Sim, M., Patel, A., and Weber, R. "Should Optional Properties be Used in Conceptual Modeling? A Theory and Three Empirical Tests," *Information Systems Research* (12:4), 2001, pp. 384-405.

Bousfield, W. A. "The Occurrence of Clustering in the Recall of Randomly Arranged Associates," *Journal of General Psychology* (49), October 1953, pp. 229-240.

Bunge, M. *Treatise on Basic Philosophy: Volume 3: Ontology I: The Furniture of the World*, Reidel, Boston. 1088.

Denkel, A. *Object and Property*, Cambridge University Press, Cambridge, UK, 1996.

Ericcson, K.A., and Simon, H. A. *Protocol Analysis*, MIT Press, Cambridge, MA, 1984.

Gemino, A. *Empirical Methods for Comparing System Analysis Modeling Techniques*, Unpublished Ph.D. Dissertation, University of British Columbia, Vancouver, BC, 1999.

Halpin, T. A. *Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design*, Morgan Kaufman, San Francisco, 2001.

Hirschheim, R., Klein, H., and Lyytinen, K. *Information Systems Development and Data Modeling: Conceptual Foundations and Philosophical Foundations*, Cambridge University Press, Cambridge, UK, 1995.

Miller, G. A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity to Process Information," *Psychological Review* (63), 1956, pp. 81-97.

Moody, D. L., and Shanks, G. "Improving the Quality of Entity-Relationship Models: An Action Research Programme," *The Australian Computer Journal* (30), 1998, pp. 129-138.

Newell, A. C., and Simon, H. A. *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, NJ, 1972.

Shanks, G., Moody, E., Nuredini, D., Tobin, D., and Weber, R. "Representing Things and Properties in Conceptual Modeling: An Empirical Evaluation," in *Proceedings of the Eleventh European Conference on Information Systems*, Naples, Italy, 2003.

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., and Weber, R. "Representing Part-Whole Relationships in Conceptual Modeling: An Empirical Evaluation," in *Proceedings of the Twenty-Third International Conference on Information Systems*, L. Applegate, R. Galliers, and J. I. DeGross (eds.), Barcelona, 2002, pp. 89-110.

Sia, S. K., and Soh, C. "Severity Assessment of ERP-Organizational Misalignment: Honing in on Ontological Structure and Context Specificity," in *Proceedings of the Twenty-Third International Conference on Information Systems*, L. Applegate, R. Galliers, and J. I. DeGross (eds.), Barcelona, 2002, pp. 723-729.

Simsion, G., and Witt, G. *Data Modeling Essentials: Analysis, Design and Innovation* (2nd ed.), The Coriolis Group, Scottsdale, AZ, 2001.